1

Which of the following is NOT a guideline for establishing causality?

- Look for cases where correlation remains while other factors vary.
- Check if the effect is present or absent when the explanatory variable is present or absent.
- Perform a randomized, controlled experiment.
- Check if the effect is present or absent when the response variable is present or absent.

RATIONALE

We don't need to check if the effect is present or absent with the response, but the explanatory variable. It might be the case the the explanatory variable has many effects.

CONCEPT

Establishing Causality

2

This scatterplot shows the performance of a pressure sensor using two variables, pressure and voltage

The equation for the least-squares regression line to this data set is

The predicted value for the voltage for a pressure of 50 MPa is _____.

- 2580 mV
- 2582 mV
- 2502 mV
- 2560 mV

RATIONALE

In order to get the predicted voltage when the pressure is 50 MPA, we simply substitute the value 50 in our equation for x. So we can note that:

CONCEPT

Predictions from Best-Fit Lines

3

A clinic has recorded the age, x, versus weight, y, of many babies for their first 12 months of life, and claim the line of best fit is $\hat{y} = 0.60x + 3.3$, where y is in kg, and x is in months.

A new baby, who is 10 months and weighs 10 kg, is added to the clinic records.

What is the residual of the data for this new baby?

- 0.4 kg
- –0.4 kg
- 0.7 kg
- –0.7 kg

RATIONALE

Recall that to get the residual, we take the actual value - predicted value. So if the actual age of the baby is 10 kg and the resulting actual weight 10 kg, we simply need the predicted weight. Using the regression line, we can say:

The predicted weight is 9.3 kg. So the residual is:

CONCEPT

Residuals

4

Using the provided scatterplot, select the correct direction of the blue outlier.

- The outlier is in the y-direction.
- The outlier is in the x-direction.
- The outlier is in both the x- and y- direction.
- The outlier is in neither the x- nor y- direction.

RATIONALE

The direction of the outlier is based on its location. The outlier in this graph is in the direction of y, since it is far away from the scatterplot in the vertical measures.

CONCEPT

Outliers and Influential Points

5

Which of the following statements is true?

- Only a correlation equal to 1 implies causation.
- A correlation equal to 1 or -1 implies causation.
- High correlation does not necessarily imply causation.
- Only a correlation equal to -1 implies causation.

RATIONALE

Correlation measures the strength and direction of linear association between 2 variables. You have to be careful, however; just because you find an association does not mean the change in one variable causes the other. A stricter set of conditions is required for causality.

CONCEPT

Correlation and Causation

6

This scatterplot shows the performance of a pressure sensor using two variables, pressure and voltage.

Select the answer choice that accurately describes the data's form, direction, and strength in the scatterplot.

- Form: Non-Linear
  Direction: Negative
  Strength: Strong
- Form: Linear

  Direction: Negative

  Strength: Weak
- Form: Non-Linear
  Direction: Positive
  Strength: Weak
- Form: Linear
  Direction: Positive
  Strength: Moderate

RATIONALE

If we look at the data, there is a lot of variation/scatter which means there is a weak or moderate relationship. As pressure goes up, the voltage goes down, so the direction is negative. Finally, there appears to be a relatively linear form since a straight line would capture the data fairly well.

CONCEPT

Describing Scatterplots

7

You skipped this question and it was marked incorrect.

For the data plotted in the scatterplot below, the value was calculated to be 0.5152.

Which of the following sets of statements is true?

- 26.5% of the variation in pressure can be explained by the voltage. The correlation coefficient, r, is -0.265
- 51.5% of the variation in voltage can be explained by the pressure. The correlation coefficient, r, is -0.718
- 71.8% of the variation in pressure can be explained by the voltage. The correlation coefficient, r, is -0.718
- 48.5% of the variation in voltage can be explained by the pressure. The correlation coefficient, r, is -0.265

RATIONALE

The coefficient of determination measures the percent of variation in the outcome, y, explained by the regression. So a value of 0.515 tells us the regression with pressure, x, can explain 51.5% of the variation in voltage, y.

We can also note that
Note that the sign on correlation is the direction of the scatterplot so it is -0.718.

CONCEPT

Coefficient of Determination/r^2

8

You skipped this question and it was marked incorrect.

For a Biology assignment, Lisa collected data on plant growth of a sunflower every week for 9 weeks. When Lisa first planted the sunflower, it was 10 centimeters tall. The time (in weeks) is plotted against the height (in centimeters) as shown below.

Using the best-fit line, approximately how tall was the sunflower plant during the fifth week?

- 34 centimeters
- 33 centimeters
- 35 centimeters

- 30 centimeters

RATIONALE

To get a rough estimate of the height at Week 5, we go to that point on the horizontal axis and then see where it falls on the best-fit line. This looks to be about 30 cm.

CONCEPT

Best-Fit Line and Regression Line

9

You skipped this question and it was marked incorrect.

Which statement is true regarding correlation?

- Correlation can only be positive.
- Correlation can only be negative.
- Correlation can be used to determine the direction of the relationship between two variables.
- Correlation is a quantitative measure of the form between two variables, as seen on a scatterplot.

RATIONALE

We note that correlation is a measure of the strength and the direction of the linear association between two quantitative variables.

CONCEPT

Correlation

10

You skipped this question and it was marked incorrect.

Sam rolls two dice, one labeled “x” and the other “y.” He rolls each of the dice six times and records the (x, y) measurements as follows:

	"x" die	"y" die
Roll 1	1	4
Roll 2	2	3
Roll 3	2	2
Roll 4	5	2
Roll 5	4	4
Roll 6	6	5

For the "x" die, the mean is 3.3 and the standard deviation is 2.0.
For the "y" die, the mean is 3.3 and the standard deviation is 1.2.

Using the formula below or Excel, find the correlation coefficient, r, for this set of outcomes Sam rolled. Answer choices are rounded to the nearest hundredth.

- 0.81
- 0.23
- 0.28
- 0.82

RATIONALE

In order to get the correlation, we can use the formula
Correlation can be quickly calculated by using Excel. Enter the values and use the function "=CORREL(".

CONCEPT

Correlation

11

You skipped this question and it was marked incorrect.

Stacey finds a scatterplot that shows data for nine schools. It relates the percentage of students receiving free lunches to the percentage of students wearing a bicycle helmet. The plot shows a strong negative correlation.

Stacey recalls that correlation does not imply causation. In this example, Stacey sees that increasing the percentage of free lunches would not cause children to use their bicycle helmets less.
Identify the confounding variable that is causing Stacey's observed association.

- Parents' annual salary
- School budget
- The number of bikes at each school
- Helmet brands

RATIONALE

A confounding variable is a variable that helps to explain the correlation between 2 variables. It must be related to both variables. We can note that parents' salary would determine if a student qualifies for free school lunches. The higher the salary, the lower percentage of free lunches. We

can also note that as a parent's salary increases, bicycle helmet use should increase as they would be able to afford helmets. So, this confounding relationship helps to explain the reason we see this association. It is not the case that helmet use and receiving free lunches has any type of causal relationship.

CONCEPT

Correlation and Causation

12

You skipped this question and it was marked incorrect.

Data for weight (in kilograms) and height (in inches) of babies is entered into a statistics software package and results in a regression equation of $\hat{y} = 1.2x - 20.7$.

What is the correct interpretation of the slope if the weight is the response variable and the height is the explanatory variable?

- The weight of a baby decreases by 20.7 kilograms, on average, when the baby's height increases by 1 inch.
- The weight of a baby increases by 20.7 kilograms, on average, when the baby's height increases by 1 inch.
- The weight of a baby increases by 1.2 kilograms, on average, when the baby's height increases by 1 inch.
- The weight of a baby decreases by 1.2 kilograms, on average, when the baby's height increases by 1 inch.

RATIONALE

When interpreting the linear slope we generally substitute in a value of 1. So we can note that in general, as x increases by 1 unit, the slope tells us how the outcome changes. So for this equation, we can note that as x (height) increases by 1 inch, the outcome (weight) will increase by 1.2 kg on average.

CONCEPT

Interpreting Intercept and Slope

13

You skipped this question and it was marked incorrect.

Gary Sandoval is a photographer who is wondering if there is an association between the number of photographs he takes and percent cloud coverage. His record is shown in the scatterplot.

How many photographs did he take when the cloud coverage was 10 percent or more?

- 300
- 550
- 450
- 750

RATIONALE

In order to find the total number greater than 10%, we must add all the values 10% and above.

At 10%, there were 300 photographs.
At 11%, there were 200 photographs.
At 12%, there were 250 photographs.

So the total is 300 + 200 + 250 = 750 photographs.

CONCEPT

Scatterplot

14

You skipped this question and it was marked incorrect.

A correlation coefficient between number of miles driven and number of gallons of gas used is most likely to be _____.

- between 0 and 1
- between 1 and 2
- between -1 and -2
- between 0 and -1

RATIONALE

Recall that correlation must always be between -1 and 1. So we simply want to know if there is a positive or negative association. The number of miles driven would require more gallons of gas, so we expect a positive relationship and it would be between 0 and 1.

CONCEPT

Positive and Negative Correlations

15

You skipped this question and it was marked incorrect.

Jaime finished analyzing a set of data with an explanatory variable x and a response variable y.

He finds that the mean and standard deviation for x are 5.43 and 1.12, respectively. The mean and standard deviation for y are 10.32 and 2.69, respectively.

The correlation was found to be 0.893.

Select the correct slope and y-intercept for the least-squares line.

- Slope = -2.14
  y-intercept = -3.03
- Slope = 0.37
  y-intercept = -1.33
- Slope = 2.14
  y-intercept = -1.30
- Slope = -0.37
  y-intercept = -3.03

RATIONALE

We first want to get the slope. We can use the formula:
To then get the intercept, we can solve for the y-intercept by using the following formula:

We know the slope, , and we can use the mean of x and the mean of y for the variables and to solve for the y-intercept, .

CONCEPT

Finding the Least-Squares Line

16

You skipped this question and it was marked incorrect.

For what reason may the correlation in this scatterplot be affected?

- It may be affected by an influential point.
- It may be affected by inappropriate grouping.

- It may be affected by non-linearity.
- It is impossible to determine.

RATIONALE

Recall that correlation measures linear association, and this graph is not linear.

CONCEPT

Cautions about Correlation

17

You skipped this question and it was marked incorrect.

This scatterplot shows the performance of a pressure sensor using two variables, pressure and voltage

Which answer choice correctly indicates the explanatory variable and the response variable of the scatterplot?

- Explanatory variable: Voltage
  Response variable: Pressure Sensor
- Explanatory variable: Pressure
  Response variable: Pressure Sensor
- Explanatory variable: Pressure
  Response variable: Voltage
- Explanatory variable: Voltage
  Response variable: Pressure

RATIONALE

The variable on the vertical axis is the outcome or response, while the horizontal axis is the explanatory variable. So we can note voltage is response and pressure is explanatory.

CONCEPT

Explanatory and Response Variables

18

You skipped this question and it was marked incorrect.

James takes two data points from the weight and feed cost data set to calculate a slope, or average rate of change. A hamster weighs 3 pounds and costs $3.50 per week to feed, while a Chihuahua weighs 4.8 pounds and costs $6.20 per week to feed.

Using weight as the explanatory variable, what is the slope of the line between these two points? Answer choices are rounded to the nearest hundredth.

- $2.80 / lb.
- 1.50 / lb.
- $0.67 / lb.
- $0.36 / lb.

RATIONALE

In order to get slope, we can use the formula: . Using the information provided, the two points are: (3 lb., $3.50) and (4.8 lb., $6.20). We can note that:

CONCEPT